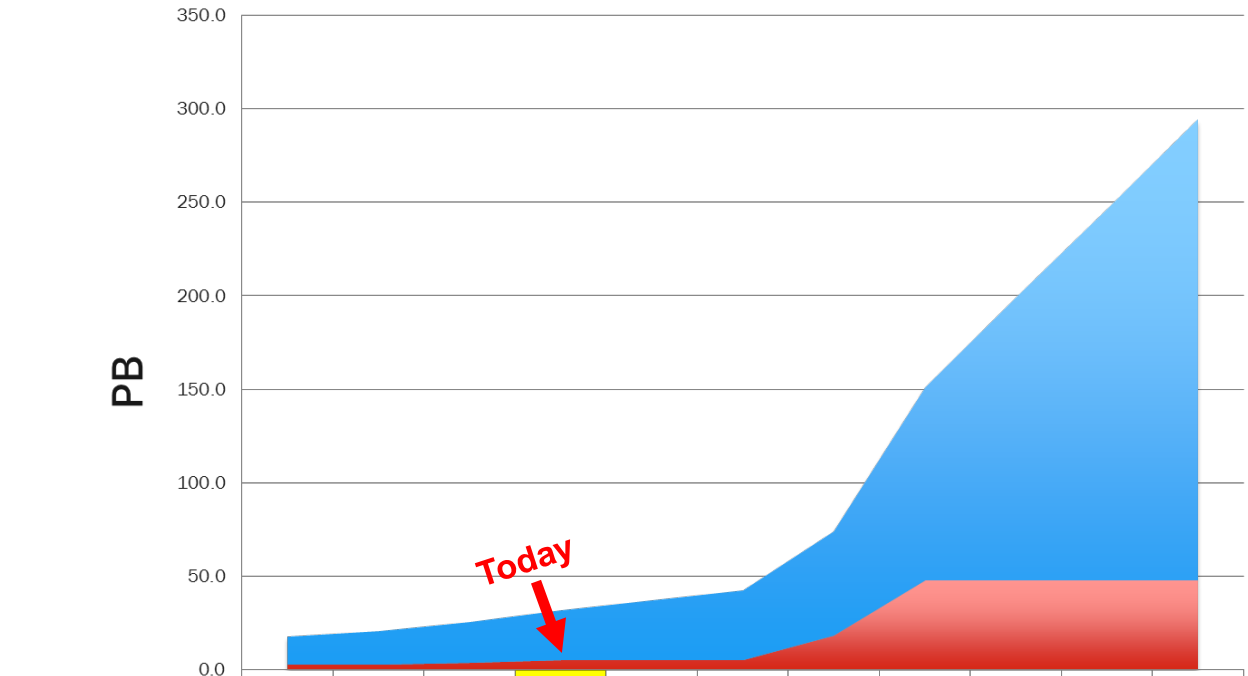
# GHRC Cloud Effort

Will Ellett

2018 GHRC User Working Group Meeting
November 13-14, 2018

NASA
itsc
INFORMATION TECHNOLOGY AND SYSTEMS CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

# Why Cloud?




| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cumulative Archive Size (PB) | 15.0 | 17.7 | 21.6 | 26.8 | 32.0 | 37.2 | 55.6 | 103.4 | 151.1 | 198.9 | 246.6 |
| Archive Growth Rate (PB) | 2.6 | 2.8 | 3.9 | 5.2 | 5.2 | 5.2 | 18.4 | 47.7 | 47.7 | 47.7 | 47.7 |

# Cumulus: What is it?

## What is Cumulus?

- Lightweight, cloud optimized framework for data ingest, archive, distribution and management

## Goals - Provide DAAC functionality

- Data Ingest & Processing
- Archiving & Distribution
- Metadata & Metrics
- DAAC-specific customizations
- DAACs can create re-usable, compatible containers to share with other DAACs

# GHRC Cloud Migration Background

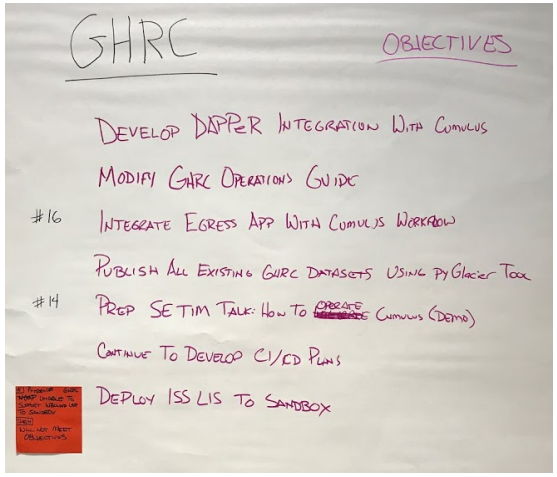
**Cumulus Prototype**

- Cumulus Prototype development began in October 2016
- Led by **Rahul Ramachandran** & **Katie Baynes**
- Exploratory process to help DevSeed to better understand DAAC requirements
- **Shannon Flynn** represented GHRC, supporting DevSeed throughout the prototype period
- Prototype used six GHRC datasets to test initial ingest, processing, archive and distribution workflows
- GHRC provided findings & recommendations to ESDIS at the end of the prototype evaluation period

# GHRC Cloud Migration Background

## GHRC Cloud Migration

- GHRC Cloud Migration Effort began in November 2017
- Collaborated with the Cumulus core development team
- Introduced Agile project management to GHRC
- Attended quarterly planning at Goddard Space Flight Center
  - 2 day meeting with Cumulus core and other DAAC's working on cloud migration
- Daily 15 minute meetings



# Training and Development Improvements

**Amazon Web Services Training**

- GHRC researched several AWS training providers
- Choose QWIKLABS based on recommendation from LP DAAC
- Hands-on training was tailored to specific roles: Development, IT & Operations
- Provided feedback outlining our experiences provided to ESDIS

# Cloud Development

**Datasets**

- Focused on learning Cumulus
- Developed sample workflows with GHRC datasets
- **Abdelhak Marouane** contributed fixes and updates included to Cumulus software
- Prototyped Near Real-Time (NRT) & On-going dataset deployments in Cumulus
- Updated Common Metadata Repository (CMR) & Earthdata Search (test environment)
- Ran ISS LIS NRT for over a month in GHRC AWS environment
- Created draft of document for publishing Datasets

# Tested EGRESS App with ASF

## Egress Solution Evaluation

- ASF DAAC developed EGRESS tool to monitor and control dataflow out of the cloud
- GHRC worked with ASF to test egress
- Provided feedback on the deployment & usability of the EGRESS App

# Lessons Learned

- Quarterly planning established by ESDIS provided insights that helped improve GHRC processes
- Cloud development process improvements incorporated into all GHRC DAAC activities
- Large learning curve for Cumulus onboarding
- Current data publication may be impacted due to additional Cumulus overhead when we run operationally
- Need improved data publication tool (Earthdata Pub)

# Future Cloud Activities

- GHRC will be the first DAAC to fully operate in the cloud
- Migrate GHRC data into test environment (by JUL 2019)
- Migrate GHRC data into production environment (by SEP 2019)
- Run Parallel Operations by end of FY2019
- Begin cost modeling evaluation/validation
- Investigation of data service integration
  - Earthdata Cloud 2021 services group
  - Earthdata Search
- Deployment Procedures
  - Guide for deploying the DAAC in the cloud
  - Update automated deployment instructions
  - Update operational procedures for operating in the cloud

# Demos

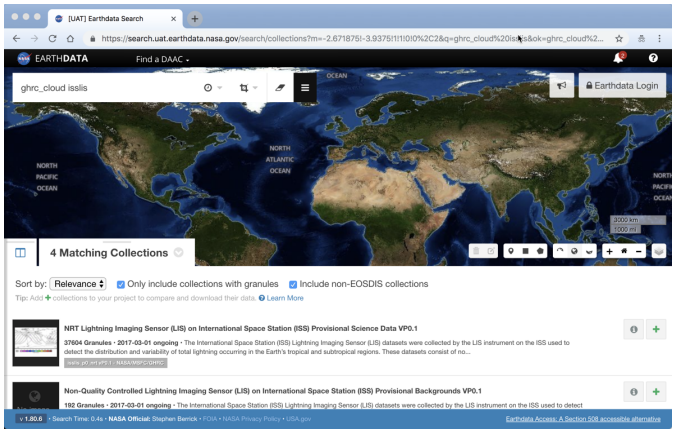
## Granule Restore

Collection list

Show 10 entries

| COLLECTION NAME | GLACIER STATUS | S3 STATUS | GRANULES COUNT ON PREM |
|---|---|---|---|
| aces1cont | Complete: 6<br>Failed: 0<br>Inprogress: 20<br>Waiting: 0<br>**Total: 26** | Complete: 6<br>Failed: 0<br>Inprogress: 0<br>Waiting: 20<br>**Total: 26** | 26 |
| aces1efm | Complete: 14<br>Failed: 0<br>Inprogress: 0<br>Waiting: 0<br>**Total: 14** | Complete: 14<br>Failed: 0<br>Inprogress: 0<br>Waiting: 0<br>**Total: 14** | 14 |
| aces1log | Complete: 27<br>Failed: 0<br>Inprogress: 0<br>Waiting: 0<br>**Total: 27** | Complete: 27<br>Failed: 0<br>Inprogress: 0<br>Waiting: 0<br>**Total: 27** | 27 |
| aces1time | Complete: 19<br>Failed: 0<br>Inprogress: 0<br>Waiting: 0<br>**Total: 19** | Complete: 19<br>Failed: 0<br>Inprogress: 0<br>Waiting: 0<br>**Total: 19** | 19 |
| aces1trig | Complete: 21<br>Failed: 0<br>Inprogress: 0<br>Waiting: 0<br>**Total: 21** | Complete: 21<br>Failed: 0<br>Inprogress: 0<br>Waiting: 0<br>**Total: 21** | 21 |

## ISS LIS



# THANK YOU!

Discussion

2018 GHRC User Working Group Meeting
November 13-14, 2018

NASA

itsc
INFORMATION TECHNOLOGY AND SYSTEMS CENTER

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE